Brazilian Journal of Information Science

LEI DE LOTKA APLICADA À PRODUÇÃO CIENTÍFICA DA ÁREA DE CIÊNCIA DA INFORMAÇÃO

Maria Isabel Martín Sobrino
Ana Isabel Pestana Caldes
António Pulgarín Guerrero

Departamento de Informação e Comunicação
Universidade de Estremadura

RESUMO

Apresentamos uma aplicação da Lei de Lotka ao conjunto de autores com publicações no campo da Ciência da Informação, entre 1996 e 2007. A aplicação realizada segue a metodologia de Lee Pao (1985). Foram selecionados todos os autores que surgiram no campo autor, não se efetuou nenhum corte na distribuição e a estimação do valor crítico calculou-se utilizando a fórmula proposta por Nicholls (1989). Os resultados mostram os seguintes dados: uma pendente igual a '-2.75', inferior à obtida tanto no trabalho de Voos (1974) como no de Sen, Taib e Hassan (1996), neste mesmo campo; uma porcentagem de autores, realizadores de um só trabalho, igual a 79% e um excelente ajuste a Lei de Lotka, ao ser aplicado o teste de Kolmogorov-Smirnov.

Palavras-chave: Lei de Lotka; Produção Científica; Ciência da Informação; Bibliometria.

INTRODUÇÃO

Em 1926, Alfred J. Lotka, examinou a distribuição de freqüências da produtividade científica de químicos e físicos (publicações listadas em *Chemical Abstracts,* desde 1907 a 1916, e em Auerbach's *Geschichtstafeln der Physik,* desde o início da sua publicação até 1900), observa-se uma relação quantitativa entre os autores e a sua produção científica. A observação de Lotka mostra uma distribuição assimétrica (como posteriormente aconteceu com a de Bradford ou Zipf), com uma concentração de artigos entre poucos autores (autores grandes produtores), enquanto que os artigos restantes estariam distribuídos entre uma grande quantidade de autores. A correlação entre autores e a sua produtividade, no caso estudado por Lotka, mostrou uma pendente negativa, próxima de -2.

Desde então, muitos têm sido os trabalhos realizados, com o objetivo de aplicar ou reformular a Lei de Lotka, obtendo-se resultados contraditórios e, nem sempre, com bons resultados (OPPENHEIMER, 1986).

Existe uma extensa literatura que trata sobre a aplicação da Lei de Lotka. Entre elas podemos destacar os seguintes trabalhos:

Murphy (1973) aplicou a Lei de Lotka no campo das humanidades, concluindo que a Lei se cumpria, não tendo aplicado nenhum teste estatístico, para comprovar o grau de significância.

Em 1974, Voos estudou a produtividade dos autores no campo da Ciência da Informação entre, 1966 e 1970, e comparou os resultados com a observação de Lotka (*n* = 2), descobriu que a distribuição de autores se ajustava muito bem a uma nova constante igual a $x^{-3,5}$. A percentagem de autores com um só trabalho, obtido por Voos, foi de 88%, ao invés de 60% obtido por Lotka. Ainda que Voos realizasse o estudo dos cinco anos separadamente, ano por ano, comprovamos que se considerarmos o conjunto dos cinco anos também se ajusta.

Schorr publicou três artigos nos quais apresentava outras aplicações da Lei de Lotka: as ciências das bibliotecas, a organização de bibliotecas e a história da medicina legal. No seu primeiro artigo (SCHORR, 1974) encontrou uma Lei que era quádruplo $\left(x^{-4}\right)$, em vez da quadrática inversa de Lotka $\left(x^{-2}\right)$. Em outras experiências sobre organização de bibliotecas (SCHORR, 1975a), depois de aplicar o teste χ^2, concluiu que esta disciplina se ajustava à Lei de Lotka. No seu terceiro artigo (SCHORR, 1975b), estudou a produtividade na história da medicina legal e aplicando o teste de χ^2, descobre que os autores com múltiplos trabalhos estavam muito abaixo do esperado segundo a Lei de Lotka (<60 %), concluindo que esta Lei não era a mais correta para esta matéria.

Coile, num artigo publicado em 1977, nega a conclusão do segundo artigo de Schorr, sobre organização de bibliotecas, afirmando que não estava correto já que aplicou em alguns dados, um teste estatístico inapropriado (referindo-se ao teste de χ^2). Coile, depois de apresentar a Lei de Lotka extraída do trabalho original, examinou e comprovou os dados do artigo de Murphy, em humanidades, e os de Schorr, em organização de bibliotecas, utilizando o teste de Kolmogorov-Smirnov (K-S), concluindo que em nenhum dos casos se cumpria a Lei de Lotka.

Dois anos depois, Radhakrishnan e Kerdizan (1979) verificaram que a Lei de Lotka não se aplicava corretamente aos dados sobre publicações na área de Informática, observando que estava mais próximo de uma lei x^{-3}. Estes autores assumiram que quando um trabalho tinha vários autores, a cada um dos autores pertencia o trabalho completo (*normal count*). Esta associação tinha um efeito incontestável de valorizar o número de autores que escreviam apenas um trabalho, e eram da opinião de que só se registrava o artigo ao autor principal ou ao primeiro autor, como fez Lotka "*straight count*", ajustando-se a uma quadrática inversa (x^{-2}). Para provar esta hipótese, examinaram uma amostra aleatória deste campo, registrando apenas um autor para cada trabalho, e sem aplicar nenhum teste estatístico, concluíram que os dados se ajustavam à Lei de Lotka. Em seguida, realizaram a mesma experiência com os dados do primeiro artigo de Schorr, sobre ciências das bibliotecas, registrando apenas o primeiro autor e sem aplicar nenhum teste, obtiveram resultados que se ajustava a uma lei x^{-3}, em vez de x^{-4}, obtida por Schorr.

Vlachý (1978), na secção referente à bibliografia do primeiro número de *Scientometrics*, apresenta uma bibliografia sobre Lotka e trabalhos relacionados, entre eles sobre Bradford e Zipf, assim como distribuição de freqüências e de bibliometria. Em trabalhos anteriores Vlachy (1974, 1976) tinha encontrado discrepâncias entre os dados empíricos e a lei quadrada inversa, isto é, o valor do expoente da Lei de Lotka era variável.

Em 1985, Miranda Lee Pao publica um artigo em que apresenta o processo da aplicação da Lei de Lotka, passo a passo, calculando os valores da constante e expoente, baseando-se no método de Lotka, assim como na utilização de um teste para comprovar o grau de significância. Um ano depois (PAO, 1986), esta mesma autora aplica este procedimento a 48 conjuntos de autores, representando 20 campos científicos distintos. Os resultados são conclusivos, em 80 % dos casos ajusta-se a lei de Lotka.

Duas modificações ao procedimento de Pao são propostas por Nicholls (1986) e aplicadas a 15 amostras de humanidades, Ciências Sociais e Ciências. As modificações referem-se ao cálculo da pendente (expoente), que propõe que se calcule por aproximação da máxima probabilidade (métodos iterativos numéricos) e de forma a considerar todos os co-autores dos trabalhos. Para o cálculo do valor crítico, que servirá de comparação com a diferença máxima (Dmax), propõe a seguinte fórmula:

$$v.c. = 1{,}63 / \left(\sum y_x + \left(\sum y_x / 10 \right)^{1/2} \right)^{1/2}$$

Nicholls (1989), num segundo trabalho, opina que existe uma literatura considerável sobre a válida empírica da Lei de Lotka, não obstante estes estudos são na sua maioria incomparáveis e não conclusivos, tendo diferenças substanciais no método aplicado. De acordo com Nicholls, os principais elementos implicados no acerto dos dados empíricos a um modelo bibliométrico são: a especificação do modelo, a medida das variáveis, a organização dos dados, a estimação dos parâmetros e o cálculo do grau de significância.

Gupta (1987), num estudo sobre entomologia da Nigéria, analisa e estuda modelos de produtividade dos autores e comprova a aplicabilidade da Lei de Lotka para quatro conjuntos diferentes de dados. Mostra que a Lei de Lotka, na sua forma original, como quadrática inversa não é aplicável a nenhum dos quatro conjuntos de dados. Num outro trabalho posterior, Gupta (1989a) aplicou a Lei de Lotka à literatura sobre psicologia na África. Observa que a Lei não era aplicável aos dados na sua forma generalizada (n = 2,8), aplicando neste caso ambos os testes de estatística (K-S y χ^2).

BJIS, v.2, n.1, p.16-32, jan./jun. 2008. Disponível em: <http://www.bjis.unesp.br/>. ISSN: 1981-1640

Num terceiro trabalho, no campo da bioquímica da Nigéria, este mesmo autor (GUPTA, 1989b), criou quatro diferentes fichários, um com todos os autores, outro com apenas os primeiros, com os não colaboradores e um quarto fichário só com os co-autores, verificou que se podia aplicar a Lei de Lotka nos quatro casos, mas com valores distintos no expoente. Para comprovar o ajuste utilizou o teste Kolmogorov-Smirnov, a um nível de significância de 0,01.

Sen, Taib e Hassan (1996), trabalhando no domínio das Ciências da Informação, tentam validar a Lei de Lotka, verificando que esta é aplicável a esse campo.

Jiménez Contreras Moya Anegón (1997), analisa a produtividade dos autores no campo da Biblioteconomia e Documentação na Espanha, concluindo que a Lei de Lotka descrevia muito bem a distribuição dos dados.

Pulgarín e Gil-Leiva (2004) desenvolvem um estudo com referências sobre indexação desde 1956 a 2000, concluindo que os dados se ajustam a uma distribuição de Lotka.

Urbizagástegui (2006), recentemente, analisou a distribuição da potência inversa, e descreveu passo a passo a aplicação do modelo proposto por Pao em 1985. A literatura estudada ajusta-se ao modelo de Lotka.

2 METODOLOGÍA

Desde 1996, ano em que Sen, Taib e Hassan publicaram um trabalho sobre isso, não foi observado nenhum trabalho posterior em que se tenha aplicado a Lei de Lotka, no campo das Ciências da Informação. Este foi o motivo que nos levou a atualizar a aplicação da Lei de Lotka, seguindo a metodologia de Pao (1985), para este campo.

Os dados foram obtidos a partir da base de dados *Library and Information Science Abstracts* (LISA), realizando uma pesquisa retrospectiva desde 1996 a princípios de 2008, utilizando o termo “*Information Science*” como descritor.

A contagem foi efetuada, atribuindo os mesmos créditos a cada um dos autores que apareciam em cada trabalho (*Normal count*) (NICHOLLS, 1986; LINDSEY, 1982).

Para o cálculo da pendente não se efetuou nenhum corte, quer dizer, procedeu-se à determinação desse parâmetro utilizando todos os dados. A partir desta perspectiva, menos preceptiva e mais descritiva, eliminar dado é uma perda objetiva de informação e tem algo de artefato científico.

A Lei de Lotka estabelece que o número de autores, y_x, cada um deles com '*x*' trabalhos é inversamente proporcional a *x*, que é a produtividade de cada autor individual.

A relação se expressa como $x^n \cdot y_x = c$; x = 1, 2,.., x_{max}, *c* > 0, *n* > 1,

onde y_x representa a probabilidade de um autor publicar '*x*' vezes nessa área, x_{max} representa o máximo valor de produtividade, e '*n*' e '*c*' são dois parâmetros que é necessário estimar para cada conjunto específico de dados.

A pendente foi calculada, seguindo o protocolo proposto por Lee Pao, isto é, por método dos mínimos quadrados.

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2},$$

onde

N = Número de pares de dados considerados

X = Logaritmo decimal de *x*

Y = Logaritmo decimal de *y*

A estimação do parâmetro '*c*', percentagem de autores com um só trabalho, é mais problemático. A solução mais simples é aceitar a conclusão de Lotka que diz: *"A proporção de autores com um só trabalho é de 60 %"*, que foi a % que obteve nas suas duas amostras $6/\pi^2$. Muitos investigadores escolhem esta lei quadrática inversa para efetuar verificações porque é a mais fácil de calcular.

Extrapolando o cálculo de Lotka, para o caso especial de *n* = 2, a fórmula geral de *n* é da seguinte forma:

BJIS, v.2, n.1, p.16-32, jan./jun. 2008. Disponível em: <http://www.bjis.unesp.br/>. ISSN: 1981-1640

$$y_1 = C\left(1/1^2 \right)$$
$$y_2 = C\left(1/2^2 \right)$$
"
"
"
$$y_x = C\left(1/x^n \right)$$

Somando ambos os termos das equações obtemos:

$\sum y_x = C\left(1/1^2 + 1/2^2 + 1/3^2 + \ldots\ldots\ldots + 1/x^2 \right)$, e dividindo ambos os termos pelo número total de autores, $\sum y_x$ temos: $\frac{\sum y_x}{\sum y_x} = \left(\frac{C}{\sum y_x} \right)\left(\sum \frac{1}{x^2} \right)$.

Como $\frac{\sum y_x}{\sum y_x} = 1$, fazendo $\left(\frac{C}{\sum y_x} \right) = c$, temos que $1 = c\left(\sum \frac{1}{x^2} \right)$, e assim $c = \frac{1}{\left(\sum 1/x^2 \right)} = \frac{1}{\frac{\pi^2}{6}} = \frac{6}{\pi^2} = 0{,}6079$, já que para $n = 2$, a série $\sum 1/x^2$ converge até $\pi^2/6$.

Para o caso de outros valores fracionados, não negativos, de n, a soma da série infinita, na sua forma geral, $\sum 1/x^n$, só pode ser aproximada a uma função que calcula a soma dos primeiros P termos. O cálculo dos P = 20 primeiros termos, ignorando o cálculo do resto dos termos até ∞, encontra-se desenvolvido no trabalho de Pao (1985).

O resultado da soma desta série infinita $1 + 1/2^n + 1/3^n + \ldots + 1/x^n$, quando n > 1, para os P primeiros termos é:

$$\sum_{x=1}^{\infty} 1/x^n \cong \left[\sum_{x=1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)\left(P^{n-1}\right)} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{n+1}} \right]$$

Logo para estimar *c*, fração de autores com um só trabalho numa distribuição de autores, utiliza-se a função inversa *z* de Riemann:

$$c = \frac{1}{\sum_{x=1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)P^{n-1}} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{n+1}}}$$

Finalmente, devemos escolher um teste estatístico apropriado para verificar a significância do grau de ajuste, com um nível de importância determinado, para comprovar se a distribuição observada está conforme ou se ajusta à função de distribuição teórica.

Coile (1977) criticou o uso do teste de χ^2 por parte de alguns autores, dizendo que o valor desse teste culmina na necessidade de combinar os dados em várias categorias, sugerindo o teste de Kolmogorov-Smirnov como mais poderoso estatisticamente. Por esse motivo, este será o teste utilizado neste estudo.

3 RESULTADOS

O resultado da pesquisa efetuado na base de dados LISA, utilizando o termo "*Information Science*" como descritor, gerou um total de 2825 registros. De cada registro obtido selecionou-se o campo autor, obtendo-se 2695 autores. Os resultados da pesquisa são mostrados na Tabela 1, que indica o número de trabalhos publicados (coluna 1), o número de autores com *x* trabalhos publicados (coluna 2) e os cálculos pertinentes para poder calcular a pendente da distribuição de autores (colunas 3 a 6).

Tabela 1: Dados Observados e Dados para Calcular a Pendente

x	y	X = log x	Y = log y	XY	XX
1	2137	0	3,32980	0	0
2	341	0,30103	2,53275	0,76243	0,09061
3	104	0,47712	2,01703	0,96236	0,22764
4	48	0,60205	1,68124	1,01220	0,36247
5	27	0,69897	1,43136	1,00048	0,48855
6	12	0,77815	1,07918	0,83976	0,60551
7	9	0,84509	0,95424	0,80642	0,71419

8	4	0,90308	0,60205	0,54371	0,81557
9	2	0,95424	0,30103	0,28725	0,91057
10	2	1	0,30103	0,30103	1
11	2	1,04139	0,30103	0,31349	1,08449
12	2	1,07918	0,30103	0,32486	1,16463
13	1	1,11394	0	0	1,24086
14	1	1,14612	0	0	1,31360
15	1	1,17609	0	0	1,38319
16	1	1,20411	0	0	1,44990
25	1	1,39794	0	0	1,95423
TOTAL	**2695**	**14,7185**	**14,8318**	**7,1540**	**14,8061**

Fonte: LISA – 1996/2008.

Com os dados da tabela 1 procede-se ao cálculo da pendente (*n*).

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2} = \frac{17 \times 7{,}154 - 14{,}718504 \times 14{,}8318}{17 \times 14{,}8061 - 14{,}7185^2} = -2{,}7569$$

Para estimar *c*, procede-se ao cálculo da função z de Riemann:

$$c = \frac{1}{\sum_{x=1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)P^{n-1}} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{n+1}}}$$

Previamente se obtém a soma da série infinita para os P-1 primeiros termos:

$$\sum_{x=1}^{\infty} 1/x^{2,7569} \cong \left[\sum_{x=1}^{19} \frac{1}{x^{2,7569}} + \frac{1}{(1,57)(20)^{1,7569}} + \frac{1}{2(20^{2,7569})} + \frac{2,7569}{24(19)^{3,7569}}\right] = 1,2583$$

$$C = \frac{1}{1{,}2583} = 0{,}794723 .$$

Esta é a percentagem de autores com um só trabalho publicado na distribuição de autores, o primeiro dado da coluna 5 da Tabela 2 é a partir do qual se calculam os restantes valores teóricos.

A Tabela 2 é construída com o propósito de submeter os dados observados a um teste estatístico, para verificar o grau de significância. Neste caso aplicaremos o teste K-S.

Tabela 2: Dados para Aplicar o Teste de Kolmogorov-Smirnov.

x	y	y/∑y	∑(y/∑y)	fe	∑fe	Dmax
1	2137	0,79294	0,79294	0,79472	0,79472	0,00177
2	341	0,12653	0,91948	0,11757	0,91229	0,00718
3	104	0,03858	0,95807	0,03844	0,95074	0,00733
4	48	0,01781	0,97588	0,01739	0,96813	0,00774
5	27	0,01001	0,98589	0,00940	0,97753	**0,00836**
6	12	0,00445	0,99035	0,00568	0,98322	0,00712
7	9	0,00333	0,99369	0,00371	0,98694	0,00674
8	4	0,00148	0,99517	0,00257	0,98951	0,00566
9	2	0,00074	0,99591	0,00186	0,99137	0,00454
10	2	0,00074	0,99666	0,00139	0,99276	0,00389
11	2	0,00074	0,99740	0,00107	0,99383	0,00356
12	2	0,00074	0,99814	0,00084	0,99467	0,00346
13	1	0,00037	0,99851	0,00067	0,99535	0,00316
14	1	0,00037	0,99888	0,00055	0,99590	0,00298
15	1	0,00037	0,99925	0,00045	0,99635	0,00290
16	1	0,00037	0,99962	0,00038	0,99673	0,00289
25	1	0,00037	1	0,00032	0,99706	0,00293
TOTAL	**2695**					

Fonte: LISA -1996/2008.

A Dmax obtida ao efetuar as diferenças em valor absoluto entre as duas colunas de dados acumulados (observados e teóricos) é igual a 0,008363.

O valor crítico, a um nível de significação de $\alpha = 0{,}01$, segundo Nicholls é:

$$v.c. = \frac{1{,}63}{\sqrt{\sum y_x + \sqrt{\frac{\sum y_x}{10}}}} = \frac{1{,}63}{\sqrt{2695 + \sqrt{\frac{2695}{10}}}} = 0{,}0313$$

CONSIDERAÇÕES FINAIS

As diferenças substanciais, observadas nos diferentes estudos realizados para validar a Lei de Lotka nas diferentes matérias, radicam na metodologia aplicada, fundamentalmente (NICHOLLS, 1989). Isso provocou que os resultados tenham sido contraditórios e, nem sempre, ajustados a uma distribuição lotkiana (OPPENHEIMER, 1986).

Como se pode confirmar, com um número considerável de casos, incluindo algum dos exemplos do próprio Lotka, a Lei não se cumpre, tendo-se obtido valores da pendente diferentes a -2.

No caso de Murphy (1973) não se submeteram os dados a um teste estatístico, com o qual é impossível afirmar que se cumpre ou não a Lei de Lotka. O estudo de Voos (1974) apresenta dois problemas: o primeiro é que se estudam os dados ano a ano, como tal o período de tempo é muito pequeno, bastante menor aos dez anos como sugere Potter (1981); o segundo problema é o teste que é utilizado (χ^2), teste inapropriado. Segundo Coile (1977), requer agrupar as categorias que apresentam freqüência inferior a 5, com a conseqüente perda de informação. Schorr (1974, 1975a, 1975b), também utiliza a χ^2, como teste estatístico, o que pode deslocar os resultados do seu estudo, como comprovou Coile ao aplicar um teste adequado, como é o de Kolmogorov-Smirnov.

Não estamos de acordo com Radhakrishnan e Kerdizan (1979), ao considerar todos os autores de um trabalho (*normal count*), em lugar de fazê-lo como fez Lotka (*straigh count*), isto é, atribuir o trabalho ao autor principal, sendo o número de autores beneficiado. Este método não teria conseqüências nos tempos de Lotka, já que a percentagem de autores que publicavam em colaboração seriam muito menor em relação aos que o fazem na atualidade (uma das características da *Big Science*) (PRICE, 1963), mas na atualidade não se pode ignorar este feito e, pelo menos, temos de ter em conta todos os co-autores. Radhakrishnan e Kerdizan, também, não aplicam nenhum teste para comprovar o ajuste dos seus dados.

BJIS, v.2, n.1, p.16-32, jan./jun. 2008. Disponível em: <http://www.bjis.unesp.br/>. ISSN: 1981-1640

De acordo com Vlachý (1974, 1976), Pao (1985, 1986) e Nicholls (1986, 1989), temos de admitir que o valor do expoente da Lei de Lotka (pendente) é variável e, portanto, a constante (número de autores com um só trabalho), também, será diferente para cada distribuição de autores (diferente a 60 % do total de autores da distribuição que propôs Lotka). Igualmente, será necessário aplicar um teste estatístico apropriado, que não distorça os dados, sobretudo dos grandes produtores, como o teste de Kolmogorov-Smirnov. Acreditamos que se deve utilizar todos os co-autores, tal e qual como indicava Nicholls e, para o cálculo do valor crítico, utilizar a proposta de Pao ou a modificada de Nicholls.

Desde 1974, ano em que Voos publicou o seu artigo, até 1996, não conseguimos observar a publicação de nenhum outro artigo, aplicando a Lei de Lotka ao campo da Ciência da Informação. Este segundo artigo foi elaborado por Sen, Taib e Hassan, utilizando o índice anual de nomes de 1992 (com um resultado de 8284 nomes) e o índice anual de autores de 1993 (com um resultado de 7664 autores), da LISA. Utilizando como constante (C) o número de autores observados com um só trabalho, descobre-se a pendente que resulta ser para os dados de 1992, igual a 3,23 e para os de 1993, igual a 3,1. Conclui-se afirmando que ambas as distribuições se ajustam à Lei de Lotka, já que os dados teóricos calculados com as pendentes encontradas se aproximam bastante aos dados reais.

Em relação ao artigo de Sen, Taib e Hassan, não estamos de acordo pelos seguintes motivos:

1. O fato de selecionar os dados de um só ano parece-nos um período de tempo demasiado curto, de acordo com Potter (1981).
2. O método seguido para o cálculo da pendente parece-nos inapropriado. Utilizando os dados observados em 1992, obtivemos uma pendente igual a 3,4 para 1992, em vez de 3,23 e de 3,2 para 1993, em vez de 3,1. O método que utilizamos foi o dos mínimos quadrados e, também, de forma gráfica.
3. Os autores com um só trabalho publicado, procedentes dos dados

observados não podem substituir aos dados teóricos e a partir dai, considerando este valor como C, calcular os diferentes valores para autores com dois trabalhos, três trabalhos etc. Temos de calcular o valor teórico de C, previamente.

4. Também observamos que não existiu a aplicação de nenhum teste estatístico que justifique ou ajuste as distribuições.
5. Por último, se considera a pendente de Lotka igual a -2, para comprovar como se ajustam os dados, teriam que ter considerado também o valor de C, como refere Lotka, isto é, $0,6097 = 6/\pi^2$. Então o valor de C seria $8284 \times 0,6097 = 5050$ autores em lugar de 7229, para o caso de 1992 e de $7601 \times 0,6097 = 4634$, para o caso de 1993. Se somarmos os autores calculados com n = 2, para 1992 obtêm-se 10580 autores, em vez de 8284, e para o caso de 1993 um total de 8484, em vez de 7601.

Os nossos resultados mostram uma nova visão, em relação à metodologia aplicada até agora para o campo da Ciência da Informação.

Estamos convencidos de que esta metodologia utilizada neste estudo, é possível obter resultados mais confiável, proporcionando um maior rigor à investigação.

Utilizamos na recontagem todos os co-autores, devido às características que apresenta a investigação atual, relativamente ao trabalho em equipe, característica a que nos referimos anteriormente quando citamos Price (1963).

Não defendemos um corte à distribuição, ainda que os dados se mostrem menos atrativos. Por isso consideramos todos os dados.

Acreditamos que é fundamental aplicar uma metodologia adequada para o cálculo dos parâmetros da Lei de Lotka, a pendente da distribuição e C, proporção teórica de autores com um só trabalho. No nosso caso consideramos a metodologia proposta por Pao, para o cálculo de ambos os parâmetros, mostrando como resultado uma pendente igual a 2,7569 e um C igual a 0,794723. Como se pode observar estes resultados são menores que nos casos de Voos e de Sen, Taib e Hassan, o que significa que o número de trabalhos pontuais (autores que num dado momento

BJIS, v.2, n.1, p.16-32, jan./jun. 2008. Disponível em: <http://www.bjis.unesp.br/>. ISSN: 1981-1640

escrevem um artigo e já não o voltam a fazer, o que indicaria uma descontinuidade na investigação) tende a diminuir. Como conseqüência disto é lógico que a pendente também tenha baixado.

Para um maior rigor e como parte da metodologia a seguir, é importante a aplicação de um teste estatístico, com o objetivo de provar a hipótese de partida e poder confirmar o ajuste ou não a uma distribuição do tipo Lotka. No nosso caso optamos pelo teste não paramétrico de Kolmogorov-Smirnov, o teste pareceu-nos, igual ao que opinou Coile, mais apropriado que o de $\chi 2$ para aplicá-lo a uma distribuição assimétrica como é a de autores.

Por último, o valor crítico obtido, segue a formulação proposta por Nicholls, a um nível de significância de $\alpha = 0{,}01$, foi de 0,0313, enquanto que o de *Dmax* encontrado era de 0,008363. Como *v.c.* > *Dmax* aceita a hipótese nula. Portanto, também neste terceiro estudo, igualmente ao que aconteceu nos dois anteriores, é necessário confirmar o ajuste da distribuição de autores, no campo da Ciência da Informação, referente à Lei de Lotka.

REFERENCIAS

COILE, R. C. Lotka's frequency distribution of scientific productivity. **Journal of the American Society for Information Science**, v.28, n.6, p.366-370, 1977.

GUPTA, D. K. Lotka's law and productivity patterns of entomological research in Nigeria for the period 1900-1973. **Scientometrics**, v.12, n.1-2, p.33-46, 1987.

GUPTA, D. K. Lotka's law and its application to author productivity distribution of psychological literature of Africa for the period 1966-1975 - part II. **Herald of Library Science**, v.38, n.4, p.315-326, 1989a.

GUPTA, D. K. Scientometric study of biochemical literature of Nigeria, 1970-1984; application of Lotka's law and the 80/20 rule. **Scientometrics**, v.15, n.3-4, p.171-79, 1989b.

JIMÉNEZ CONTRERAS, E.; MOYA DE ANEGÓN, F. Análisis de la autoría en revistas españolas de Biblioteconomía y Documentación: 1975-1995. **Revista Española de Documentación Científica**, v.20, n.3, p.252-266, 1997.

LINDSEY, D. Futher evidence for adjusting for multiple authorship. **Scientometrics**, v.4, n.5, p.389-395, 1982.

LOTKA, A. J. The frequency distribution of scientific productivity. **Journal of the Washington Academy of Sciences**, v.16, n.12, p.317-323, 1926.

MURPHY, L. J. Lotka's law in the humanities? **Journal of the American Society for Information Science**, v.24, n.6, p.461-462, 1973.

NICHOLLS, P. T. Empirical validation of Lotka's law. **Information Processing & Management**, v.22, n.5, p.417-419, 1986.

NICHOLLS, P. T. Bibliometric modeling processes and the empirical validaty of Lotka's law. **Journal of the American Society for Information Science**, v.40, n.6, p.379-385, 1989.

OPPENHEIMER, C. The use of online database in bibliometric studies. In: INTERNATIONAL ON-LINE INFORMATION MEETING, London, 9., 1985. **Anais…** Oxford (England): Learned Information, 1986. p.355-364

PAO, M. L. Lotka's law: a testing procedure. **Information Processing & Management**, v.21, n.4, p.305-320, 1985.

PAO, M. L. An empirical examination of Lotka's law. **Journal of the American Society for Information Science**, v.37, n.1, p.26-33, 1986.

POTTER, W. G. Lotka's law revisited. **Library Trends**, v.30, n.1, p.21-39, 1981.

PRICE, D. J. de S. **Little science, big science**. New Cork: Columbia University Press, 1963.

PULGARÍN, A.; GIL LEIVA, I. Bibliometric analysis of the automatic indexing literature: 1956-2000. **Information Processing & Management**, v.40, n.2, p.365-377, 2004.

RADHAKRISHNAN, T.; KERDIZAN, R. Lotka's law and computer science literature. **Journal of the American Society for Information Science**, v.30, n.1, p.51-54, 1979.

SCHORR, A. E. Lotka's law and library science. **Reference Quarterly (RQ)**, v.14, n.1, p.32-33, 1974.

SCHORR, A. E. Lotka's law and map librarianship. **Journal of the American Society for Information Science**, v.26, n.3, p.189-190, 1975a.

SCHORR, A. E. Lotka's law and the history of legal medicine. **Research in Librarianship**, v.30, n.5, p.205-209, 1975b.

SEN, B. K.; TAIB, C. A.; HASSAN, M. F. Library and information science literature and Lotka's law. **Malasyan Journal of Library & Information Science**, v.1, n.2, p.89-93, 1996.

URBIZAGÁSTEGUI ALVARADO, R. A productividade dos autores na literatura de enfermagen: un modelo de aplicação da lei de Lotka. **Informaçao & Sociedade: Estudos**, João Pessoa, v.16, n.1, p.83-103, 2006.

VLACHÝ, J. Frequency distribution of scientific performance: A bibliography of Lotka's law and related phenomena. **Scientometrics**, v.1, n.1, p.109-130, 1978.

VLACHÝ, J. Distribution patterns in creative communities. In: WORLD CONGRESS OF SOCIOLOGY, 1974. **Anais…** Toronto, 1974.

VLACHÝ, Jan. Time factor in Lotka's law. **Probleme de Informare si Documentare**, vol. 10, n. 2, p. 44-87, 1976.

VOOS, H. Lotka and Information Science. **Journal of the American Society for Information Science**, v.25, n.4, p.270-272, 1974.

Maria Isabel Martín Sobrino
Universidad de Estremadura
Facultad de Biblioteconomía y Documentación
Plazuela
Ibn Marwan 06071 BADAJOZ
Tel.: 924286400
Fax: 924286401
mimarsob@alcazaba.unex.es

Ana Isabel Pestana Caldes
Universidad de Estremadura
Facultad de Biblioteconomía y Documentación
Plazuela
Ibn Marwan 06071 BADAJOZ
Tel.: 924286400
Fax: 924286401
Caldes.ana@gmail.com

António Pulgarín Guerrero
Universidad de Estremadura
Facultad de Biblioteconomía y Documentación
Plazuela
Ibn Marwan 06071 BADAJOZ
Tel.: 924286400
Fax: 924286401
pulgarin@unex.es